Ortiz-Herrero, L et al., 2018. Direct and indirect approaches based on paper analysis by Py-GC/MS for estimating the age of documents. Journal of analytical and applied pyrolysis, 131, pp.9–16.

Available at https://doi.org/10.1016/j.jaap.2018.02.018

Direct and indirect approaches based on paper analysis by Py-GC/MS forestimating the age of documents

L. Ortiz-Herrero[a,*], M.E. Blanco[a], C. García-Ruiz[b,c], L. Bartolomé[d]

[a] Analytical Chemistry Department, Faculty of Science and Technology, University of the Basque Country (UPV/EHU), Barrio Sarriena S/N, 48940 Leioa, Bizkaia, Spain.

[b] University Institute of Research in Police Sciences, University of Alcalá, Ctra. Madrid-Barcelona Km. 33.600, 28871 Alcalá de Henares, Madrid, Spain.

[c] Department of Analytical Chemistry, Physical Chemistry and Chemical Engineering, Edificio Polivalente de Química, University of Alcalá, Ctra. Madrid-Barcelona Km. 33.600, 28871 Alcalá de Henares, Madrid, Spain.

[d] Advances Research Facilities (SGIker), Faculty of Science and Technology, University of the Basque Country (UPV/EHU), Barrio Sarriena S/N, 48940 Leioa, Bizkaia, Spain.

*Corresponding author:
E-mail addresses: laura.ortiz@ehu.eus (L. Ortiz-Herrero), maria.blanco@iit.it (M.E. Blanco), carmen.gruiz@uah.es (C. García-Ruiz), luis.bartolome@ehu.eus (L. Bartolomé).

Abstract

The age of a relatively old document is one of the pending issues to be resolved in the field of forensic documentary examination. Although nowadays there are a variety of analytical methodologies focused in the analysis of inks for dating documents, the paper analysis has attained little attention. This work aims to develop two complementary approaches for estimating the age of documents based on paper analysis employing the pyrolysis technique coupled to gas chromatography with detection by mass spectrometry (Py-GC/MS): (i) a direct approach using the pyrolytic fingerprints and multivariate regression with artificially aged samples, and (ii) an indirect approach based on the identification of compounds characteristic of the document period. The direct approach has successfully allowed the age estimation of relatively old documents under police custody (up to 30 years of age) and the determination of a relation between the natural and the accelerated aging of paper under the used conditions. This approach is applicable to papers that have the same (or similar) composition and have been stored under comparable storage conditions. Additionally, the indirect approach is presented as an interesting perspective to ratify valuable information of the document age.

Keywords: Questioned document, Paper, Py: GC/MS, Age estimation, Multivariate regression.

1. Introduction

The analysis of documents is a matter of great interest in the forensic analysis field. From handwritten texts and signatures to printed papers, a broad range of document types may be subject to analysis (e.g. to identity certifying documents, commercial documents or industrial and intellectual property documents). These are commonly legal documents of mandatory compliance and/or with economic, administrative, labour and social implications. The wide variety and complexity of documents hinders their analysis and, therefore, several aspects of the forensic science aimed to study documents (named as documentoscopy) remain so far unsolved. The estimation of the age of a document is one of these pending issues to be resolved [1–3].

Although the body of a document may be constituted of many different materials—as for example granodiorite in the famous Rosetta Stone—for many centuries paper has been the principal component used in documentation all over the world [4]. Since the first paper made in China before the year 105 CE, composed of fragments of cloth or silk threads, different materials have been used. In 751, Arabs varied and improved the composition of paper to reduce the cost of manufacture, employing mixtures of linen, hemp and cotton. In the 10th century, a mixture of linen seeds was used in the Armenian paper [5]. In 1450, the supply problems that emerged with the invention of the printing press spurred the production of new types of paper, mainly based in the use of plants. Nowadays, paper is primarily manufactured with fibres of cellulose and hemicelluloses and wood fibres, with several impurities such as lignin, pectins, traces of resins, tannins, carbo- hydrates and waxes [2,4,6,7]. In addition, small amounts of organic and inorganic additives like inks, adhesives, fillers and bleaches are commonly added to improve the properties of the paper.

By analysing its composition, it is possible to characterize the origin of a paper through their trace elements, which can differ from one manufacturer to another [6,7]. By this way, a unique chemical finger- print can be identified for each batch or sheet of paper [4]. Therefore, paper could provide crucial information about the production of manuscripts that may not be found in any other kind of evidence [2,4,7–9].

Nevertheless, paper does not remain unchanged over time [10] as it is deteriorated by physicochemical and biological processes. The se- verity of these deterioration is influenced not only by the paper chemical composition, but also by some environmental factors involved in its storage and preservation (e.g. presence of microorganisms and contaminants, humidity, temperature and light) [4,6,8,11–14].

Furthermore, components such as inks are generally found in documents [2,3,15,16]. The forensic examination of inks is mainly focused on the analysis of any

written entry that may be added or altered [1,2,8,15,17], and its study is one of the most widespread tools for document dating [3,15]. Nevertheless, document dating is still one of the most burdensome tasks in the field of forensic documentary examination [1,3,8,15,18–20]. Ink analysis presents several difficulties, since not all the documents under investigation have ink entries and the application range of the existing methods is quite limited ($0 < x < 5$ years) [8,15,17,21]. In this study, the analysis of paper is proposed as a possible alternative to ink studies for estimating the age of relatively old documents (around 5–30 years).

Two main approaches can be applied when studying the document age of a questioned document: direct and indirect. The first perspective follows the degradation of known compounds over time, while the second one is based on the analysis of the ink or paper composition by comparison with other papers, ink strokes or inks of known age [2,3,18]. In both approaches, artificial aging treatments (using con- trolled UV irradiation, temperature, humidity and environmental contaminants) can be used to mimic the degradation processes, providing similar reference materials to naturally aged documents [22,23].

Nowadays, several analytical techniques have been applied for estimating the age of documents [2]. Because of their non-destructive nature, spectroscopic techniques are one of the most powerful tools used in this field [24]. These techniques can directly analyse the materials under investigation, without requiring the collection of samples, and even portable instruments have been developed [2,4,24]. Because of their microscopic applications, Fourier Transform Infrared Spectroscopy (FTIR) [4,25,26], Raman Spectroscopy [1,4], X-Ray Spectroscopy (XRF) [4,26] and Laser Ablation Inductively Coupled Plasma Mass Spectrometry (LA-ICP-MS) [4,19] have been commonly used. However, despite their effectiveness in obtaining preliminary information, their results often need to be confirmed or complemented with other analyses [2,4,19,24,25].

The complexity of the samples under investigation and the interest in obtaining wider information has led to the use of destructive techniques [16,18,19,24]. Gas Chromatography coupled to Mass Spectrometry (GC/MS) is widely recognized as the best approach for investigating organic materials in artworks and archaeological objects [27,28]. However, many organic materials encountered in cultural heritage or forensic area are complex polar compounds with high molecular weight and low volatility that cannot be analysed by this technique. In those cases, their analysis by GC/MS can be achieved by coupling analytical Pyrolysis (Py-GC/MS) [16,22–24,27]. Py-GC/MS has been used for the characterization and identification of organic materials in cultural heritage, art, conservation and restoration and archaeological fields with different

purposes: addressing attribution and dating issues, characterizing materials used in various stages of manufacture, establishing the various chemical and physical changes taking place in materials upon ageing and assessing the best conditions for long-term preservation and planning restoration, among others [12,13,22,23].

Py-GC/MS is a micro destructive technique that can be applied directly to a solid or liquid sample, without any prior treatment [11,27–29]. By applying this technique, chemical structures of thermal decomposition products can be identified with the aim of characterizing paper. This information makes possible to determine different organic materials such as waxes, resins, oils, binders, pigments and dyes [17,20,24,30]. The fragments observed provide a fingerprint characteristic of each particular sample, in terms of fragment nature and relative distribution [6,30]. The chemical composition of the original components of a sample can be reconstructed by performing a detailed interpretation of the chromatographic molecular profile of the thermal degradation products. However, this can be an arduous task for complex matrices [16,20,22–24,28]. Chemometric methods coupled with analytical techniques are powerful tools that are being increasingly used to overcome these difficulties. Multivariate chemometric techniques such as Principal Component Analysis (PCA), Cluster Analysis (CA) and Discriminant Analysis (DA) have been already used in the differentiation and characterization of papers and inks [31,32]. Among these, partial least squares (PLS) has particularly attracted researchers' attention for ink dating purposes [33–36], as it can be used for the correlation of spectral or chromatographic data with composition changes over time.

In this work Py-GC/MS is used for estimating the document age through two different approaches based on paper analysis: (i) a direct method using a multivariate regression of the pyrolytic profiles of synthetic samples correlated with their aging time in chamber, and (ii) an indirect method for the identification of characteristic paper components during different time periods. The applicability of these approaches was assessed by studying several office-type papers conserved under police custody registered from 1986 to 2011.

2. Experimental procedure

2.1. Analytical method

2.1.1. Pyrolysis temperature

In order to study the influence of the pyrolysis temperature in the fragmentation of the paper components, and consequently, the resolution of the obtained pyrogram, 3, 5 and 10 micro-samples were taken. In order to avoid the lack of representativeness of

the document sampled, a lower number of micro-samples was not tested. A Harris Micro-Punch was used for sampling (1.20 mm in diameter) white paper sheets (SelectOne, Australia, 100 g m^{-2}). 3, 5 and 10 micro-samples were pyrolysed at 280, 400, 600, 750 and 1000 °C and analysed by GC/MS. The representativeness of the entire sample (document) and the sensitivity of the analysis were examined.

2.1.2. Study of documents age

Sheets of three types of paper were chosen to construct the calibration model: white paper (SelectOne, Australia, 100 g m^{-2}), recycled paper (Steinbeis, Germany, 80 g m^{-2}) and notebook paper (Oxford- Optik Paper, 90 g m^{-2}, 4 mm squared).

Paper tests of 3 × 2 cm were artificially aged using xenon light based Solarbox 1500 equipment (CO.FO.ME.GRA. Italy). The spectral range between 300–850 nm was used and the irradiance, temperature and relative humidity were set at 550 W m^{-2}, 55 °C and 50%, respectively. The temperature was controlled by a BST (Black Standard Thermometer) probe.

The three different types of paper were exposed to a maximum of 475 h of aging in chamber under controlled factors. 15 tests for each kind of paper were aged following an isochronous timetable (475, 411, 387, 363, 339, 315, 282, 256, 239, 219, 194, 163, 89, 48 and 2 h).

A vial with 3 micro-samples of each aged paper was analysed by Py- GC/MS applying a pyrolysis temperature of 400 °C. A quality control sample (QC) made up of 3 micro-samples of white paper (100 g m^{-2}) was added after every 6 samples.

2.1.3. Py-GC/MS analysis

The analysis of the micro-samples was carried out in a microfurnace pyrolyser (5250 pyrolyser, CDS Analytical, USA), directly connected to the injector of a Gas Chromatograph coupled to Mass Spectrometer (5975C GC/MS system with Triple-Axis Detector, Agilent Technologies, USA). For the chromatographic separation, a ZB-WAX capillary column (30 m × 0.25 mm × 0.25 µm) was used. The carrier gas was helium (99.999%) with a constant flow of 1.7 mL min^{-1} and a pressure of 13.7 psi. The injector and transfer line temperatures were set at 300 °C. The injector worked in split mode (split ratio 1:10) and a solvent delay of 2.10 min was selected. The oven temperature program started at 40 °C, keeping this temperature for 2 min, and then increasing at 12 °C min^{-1} to 250 °C, temperature that was held for 1 min. The operation conditions for electron impact (EI) mass spectrometer were: an ionizing voltage of 70 eV and a SCAN range from *m*/*z* 40 to *m*/*z* 550. The structural identification of the fragments was performed by

comparison with the NIST 11 mass spectra library.

2.1.4. *Documents studied*

The Documentoscopy section (Madrid) of the Spanish General Commissary of Scientific Police provided 6 report documents under police custody registered in 1986, 1991, 1996, 2002, 2006 and 2011. Each of them was made up of 5 sheets stapled together at the top left, excluding the document from 1991 that consisted of only 2 stapled pages. The sheets employed were commercial white paper of 90 g m^{-2} in A4 size, typical of office stationery. It was not possible to ensure the same manufacturer for all of them.

Each report was printed in black ink on a single face. The logo of the Ministry of Interior was located in the upper left of each sheet, and on the left margin of the pages and at the bottom of the last sheet of the documents, the stamp of this department and the signatures of the section head and the experts (made with black and blue ink tools) were also contained.

The storage conditions since they were printed until analysed were similar for all the documents. After a brief period under office conditions, the reports were kept all together in classification boxes for several weeks. Afterwards, they were stored in files for about 6 years on shelves in an archive with controlled humidity, temperature and lighting factors. In 2012, documents dated from 2000 onward were kept in the archive while those from earlier dates were transferred to a basement warehouse under dark and low temperature conditions. Documents were delivered to the laboratory in transparent plastic covers.

One micro-sample from each top, middle and bottom of the pages of each document were sampled. A total of 15 micro-samples (5 sheets × 3 micro-samples per page) per document were taken, except for the one from 1991 (2 sheets × 3 micro-samples per page). They were stored and transported to the laboratory in closed glass vials until their later analysis.

A vial with 3 micro-samples of each page of document was analysed by Py-GC/MS applying a pyrolysis temperature of 400 and 600 °C, optimal in terms of chromatographic resolution. In addition, QC samples were added every 6 samples.

These documents were used to test the regression models developed and to identify the characteristic compounds of each of the dates studied.

2.2. *Data preprocessing and multivariate statistical analysis*

To examine the differences between the three types of artificially aged papers, a first exploratory non-supervised analysis of the obtained Py-GC/MS chromatographic

data was performed by means of Principal Component Analysis (PCA) using SIMCA 13.0 (Umetrics, Umeå, Sweden). The intensities of the chromatographic signals recorded at each retention time were used as X variables. The best grouping of QC samples was the criterion used for the choice of the data treatment. Data were centred before analysis and no scaling or transformation algorithm was applied.

After examining the data through PCA, a prediction model was constructed by applying Partial Least-Squares (PLS) regression. Only the retention times from 7 to 19.5 were used, since the direct injection of the volatile fragments obtained in the pyrolyser affects the pneumatic control of the chromatographic system causing lack of reproducibility in the early retention times ($2 < t_{R<} 7$). The pyrograms were treated as spectrums. A training set consisting of 45 micro-samples (3 micro-samples of white paper for each artificial aging time, n = 15) was used for model construction. The aging time in chamber (as dependent variable Y) was regressed against the intensities of the chromatographic signals of each retention time registered in the pyrograms (as independent variables X). This model was validated using a testing set of 63 micro-samples (3 micro-samples of white paper for each artificial aging time, n = 15, and 3 micro-samples for each notebook and re- cycled paper, after an accelerated aging of 475, 282 and 2 h).

Different mathematical pretreatments of the chromatographic data, including mean centering, Unit Variance (UV) and Pareto scaling, and logarithm and power transformations were tested. Spectral filters such as first and second derivative transformations, Multiplicative Scatter Correction (MSC), Savitzky-Golay, Standard Normal Variate (SNV), Row Center and Exponentially Weighted Moving Average (EWMA) and combinations thereof, were also applied in order to minimize the baseline noise and maximize the differences found in chromatographic signals.

A logarithmic transformation was applied to the Y variable, resulting in a significant decrease in Root Mean Square Error of Estimation (RMSEE) and Cross-Validation (RMSECV) values was observed.

No outliers were detected by means of Hotelling's T^2 control graphic for a 95% of confidence in any model. Although two samples (2 and 475 h) were detected on DModX graphic as possible moderate outliers, they were not excluded to avoid the lack of information loss of the extreme times.

3. Results and discussion

In the chromatograms obtained using the HP-5 capillary column several compounds eluted within the first minutes of analysis, while for the ZB-WAX column

fewer signals were detected before 7 min. The pyrolysis injection mode caused a displacement of the chromatographic signals of poorly retained compounds, therefore, the ZB-WAX capillary column was selected, increasing the chromatographic resolution and reproducibility.

No improvement on the chromatographic sensitivity was achieved with 5 or 10 micro-samples. Thus, 3 micro-samples of each aged paper and of each page of the documents were analysed.

Among the different pyrolysis temperatures studied, 280 °C resulted in a pyrogram with weak chromatographic peak intensities and few pyrolytic products. In agreement with results previously obtained [5], where temperatures higher than 300 °C are recommended, the signals obtained in this work increased when using 400 °C, and several specific compounds were formed due to the decomposition of cellulose, while at 600 °C a major number of markers and specific pyrolysis products were generated. Most of the signal intensities were enhanced at higher pyrolysis temperatures (750 and 1000 °C) but a large number of small and nonspecific products were produced causing a drastic decrease in chromatographic resolution. Based on the good chromatographic resolution and sensitivity obtained at 400 °C, this temperature was applied for carrying out the multivariate regression analysis (direct approach). In the indirect approach, both 400 °C and 600 °C temperatures were taken into account.

The reproducibility of sample collection was tested in order to ensure that all the analyses were carried out taking the same amount of micro-sample. In this test the Micro-Punch tool used extracted always an equal quantity of paper (RSD = 4% testing 10 replicas by weighing). Before performing the multivariate regression study, the pyrograms obtained from the three types of paper (white, recycled and notebook paper) were visually examined. In Fig. 1 the pyrograms of the three kinds of paper after an accelerated aging of 475, 282 and 2 h and at a pyrolysis temperature of 400 °C are shown. The main compounds of these pyrograms are listed in Table 1.

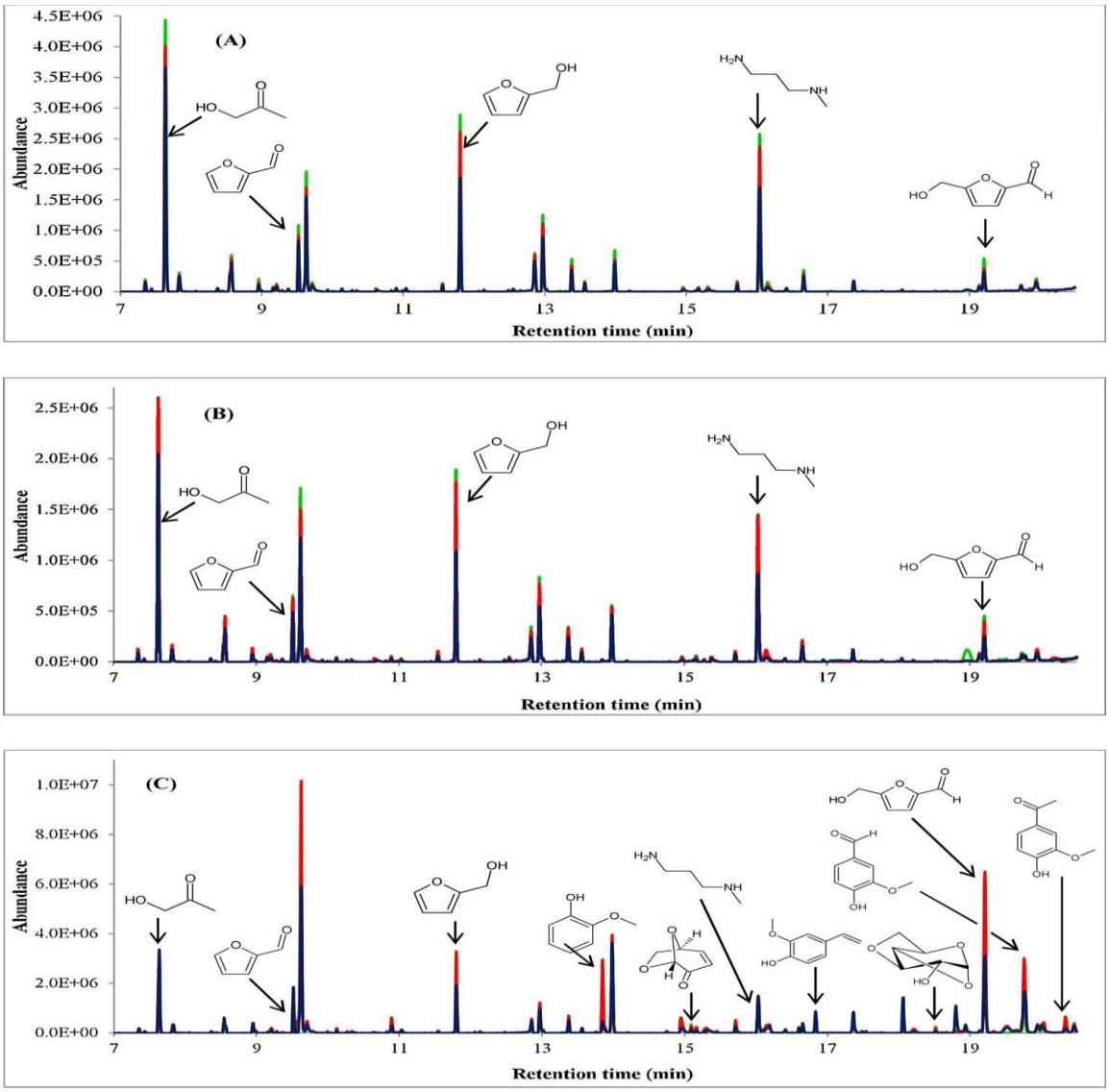

Fig. 1. Pyrolysis products identified in the pyrograms of white (A), notebook (B) and recycled (C) paper, respectively, after an accelerated aging of 475 (green line), 282 (red line) and 2 (blue line) h. Pyrolysis temperature of 400 °C. GC/MS conditions: ZB-WAX column (30 m × 0.25 mm × 0.25 μm), split injection (1:10), injection temperature of 300 °C, temperature gradient: 40 °C hold for 2 min, 12 °C min^{-1} to 250 °C, 250 °C hold for 1 min (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

The chromatographic profiles of white paper and notebook paper were practically identical. In general, the older was the paper the higher was the intensity of the chromatographic signals. This effect may indicate that the breakage of the cellulose chains of the paper made up of β-D-glucose molecules increases with age, releasing a greater number of compounds and higher chromatographic intensities. On the contrary, the pyrogram of recycled paper showed a completely different profile from the previous ones, being less evident the increment of the chromatographic signals intensity with the artificial aging. In the case of the recycled paper, it was possible to identify characteristic compounds such as 4-substituted-2-methoxyphenols formed by acid-base and redox

reactions of lignin and sugars such as levoglucosenone and 1,4:3,6- dianhydro-α-D-glucopyranose formed by the decomposition of hemi- celluloses and cellulose [37].

In the Scores Plot of the PCA model built (Fig. 2A), it was seen that the recycled paper was strongly differentiated from the white and notebook papers in the first component, in agreement with the great difference already detected by visual inspection of the chromatographic profiles. In the second component, a trend to separate the white and notebook paper samples was observed. In the Loadings Plot (Fig. 2B), it was seen that the variables X (t_R) pertaining to furfural and 5-hidroxymethylfurfural, with retention times of 9.6 and 19.20 min, respectively, were positively correlated to the first component having a great influence in the differentiation of the chromatographic intensities of the three types of paper. The chromatographic signals of these compounds were found to be more intense in the samples of the recycled paper.

The best prediction PLS models were obtained by applying SNV, EWMA and Row Center filters (Table 2), providing good precision and accuracy in prediction with 3 components (PCs), explaining an 88–92% (R^2X) and 80–93% (R^2Y) of variation in the data with predictive properties of 37–53% (Q^2), indicating that the quality of PLS regression models was acceptable with well prediction efficiency.

Table 1. Identification of pyrolysis products associated with certain chromatographic peaks detected in artificially aged white, notebook and recycled paper. Matching higher than 90% for NIST 11 mass spectra library. Analysis conditions (pyrolytic and chromatographic) as in Fig. 1.

Compound	t_R (min)	M (g/mol)	White paper	Notebook paper	Recycled paper
1-Hydroxy-2-propanone	7.65	74.08	×	×	×
Furfural	9.60	96.09	×	×	×
2-Furanmethanol	11.80	98.10	×	×	×
2-Methoxy-phenol	13.84	124.14			×
Levoglucosenone	15.10	126.11			×
N-methyl-3-propanediamine	16.04	88.15	×	×	×
2-Methoxy-4-vynilphenol	16.85	150.18			×
1,4:3,6-dianhydro-α-D-glucopyronose	18.52	144.13			×
5-Hydroxymethylfurfural	19.20	126.11	×	×	×
Vanillin (4-hydroxy-3-methoxybenzaldehyde)	19.78	152.15			×
Apocynin (1-(4-hydroxy-3-methoxyphenyl)ethanone)	20.34	166.17			×

Models were validated by means of Full Cross-Validation by the leave-one-out method. The values of Cross-Validation error RMSECV and estimation error RMSEE for the PLS model with SNV filter were 0.51 and 0.18, respectively, indicatives of the good prediction ability of the model. Likewise, the value of prediction error RMSEP was 1.11, implying that the predictive accuracy of the model was valid (see Table 2). The values of Cross-Validation, estimation and prediction errors were much larger for the PLS models in which EWMA and Row Center filters were applied.

The predictions for white paper were similar in terms of precision in the three models. In terms of accuracy, PLS model with SNV filter got a higher accuracy (relative error of 30%) comparing with the results obtained with EWMA and Row Center filters (relative error of 35%) as it can be seen in Table 3. Therefore, the application of the SNV spectral filter was optimum for the white paper (the most commonly used in questioned documents). The predictions for notebook paper instead were more accurate applying the PLS models with EWMA and Row Center filters, with identical precisions (Table 3). The recycled paper was excluded from the analysis since the differences in the chemical composition of this type of paper made it impossible to predict a date for these samples with an acceptable error (results not shown).

In the Loadings Plot applying SNV filter (Fig. 3), it has been found that the variable Y (the logarithm of the accelerated aging time) and the variables X (t_R) corresponding to 2-furanmethanol and *n*-methyl-3- propanediamine, with retention times of 11.80 and 16.04 min, respectively, had a great contribution and were positively correlated with the first component, while the compound 1-hydroxy-2-propanone, with a retention time of 7.65 min, correlates negatively with the age. This indicates that the intensities of the chromatographic signals of 2-fur- anmethanol and *n*-methyl-3-propanediamine increase with age while the intensity of 1-hydroxy-2-propanone decreases. The monitoring of the behaviour and the influence of these compounds in the aging process would be strongly recommended in further analysis.

In order to determine the correlation between artificial aging and its equivalent on real aging under police custody conditions, the PLS model using SNV filter was applied. During this study, a great variability among the different pages of the same document was observed. The abundances of some of the compounds obtained from the different pages that form the document of the same year had a dispersion measured as RSD of 21–54% for the document of 1986, 5–44% for the document of 1991, 7–16% for the document of 2006 and 17–22% for the document of 2011. This variability was not consequence of the technique applied, given that the same test performed in the controls showed a RSD of 6–8%, but of the environmental factors to which they may have been

exposed. Because of this variability, the fourth sheet of each document whose aging process was most uniform was chosen to construct the regression line (Fig. 4). The documents of 2011, 2006, 2002 and 1986 showed a good fit, while the documents of 1996 and 1991 reflected a greater deviation. This may be due to the different preservation environments that during their custody they passed through, as explained in the description of the samples. The relation between the natural aging and the accelerated aging of the papers considered was determined by means of the equation showed in Fig. 4, in which 5 h in chamber under the used conditions were equivalent to one natural year under police custody conditions. From this same regression model, other documents kept under the same conditions and made with similar class of paper could have been dated. Currently, given the impossibility of having more samples for this study from the Spanish General Commissary of Scientific Police, each of the samples used to build this regression was taken as a real sample. For this, the age of each one of the samples was estimated from the regression obtained with the dataset (n = 3), leaving out the test sample. As can be observed in Fig. 4, an acceptable estimation date was calculated (red triangle) with accuracy values between 1.8% to 7.4% in terms of RSD.

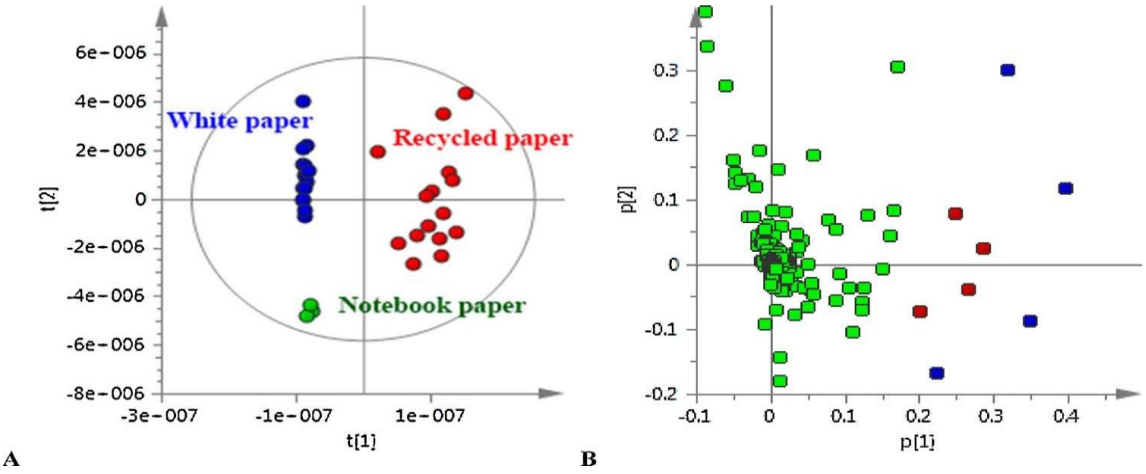

Fig. 2. A – Scores Plot of the two first components of the PCA model, with Hotelling control ellipse T^2 (95% of confidence). B – Loadings Plot of the PCA model. Points represent the variables X (t_R). Variables correspond to furfural and 5-hidrox-ymethylfurfural are coloured in blue and red, respectively. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 2. PLS model data applying centering and spectral filter transformations.

Espectral Filter	R^2X	R^2Y	Q^2	RMSEE	RMSECV	RMSEP
No spectral filter	0.92	0.80	0.37	0.30	0.47	2
First derivative transformation (15 points in filter and quadratic order polynomial fit)	0.94	0.80	0.43	0.30	0.49	1
Second derivative transformation (15 points in filter and quadratic order polynomial fit)	0.95	0.78	0.40	0.31	0.47	3
Multiplicative Scatter Correction–MSC	0.90	0.92	0.63	0.20	0.47	1
–	0.88	0.93	0.53	0.18	0.51	1
Row Center	0.92	0.81	0.38	0.30	0.47	2
Savitzky–Golay	0.94	0.80	0.40	0.30	0.49	1
Exponentially Weighted Moving Average–EWMA	0.92	0.80	0.37	0.30	0.47	2

Table 3. Predicted aging time (h) ± error for the white and notebook paper samples by applying SNV, EWMA and Row Center spectral filters.

Real aging time (h)	Predicted aging time (h)		
	SNV Filter	EWMA Filter	Row Center Filter
White paper			
48	32 ± 1	42 ± 1	41 ± 1
89	61 ± 1	38 ± 2	39 ± 2
194	285 ± 1	170 ± 1	173 ± 1
219	325 ± 1	410 ± 2	408 ± 2
239	316 ± 1	129 ± 2	133 ± 2
282	182 ± 2	119 ± 2	120 ± 2
363	367 ± 1	424 ± 1	421 ± 1
387	425 ± 1	376 ± 1	379 ± 1
475	240 ± 2	415 ± 1	408 ± 1
Notebook paper			
2	78 ± 39	13 ± 6	13 ± 6
282	1322 ± 5	203 ± 1	201 ± 1
475	2120 ± 4	427 ± 1	419 ± 1

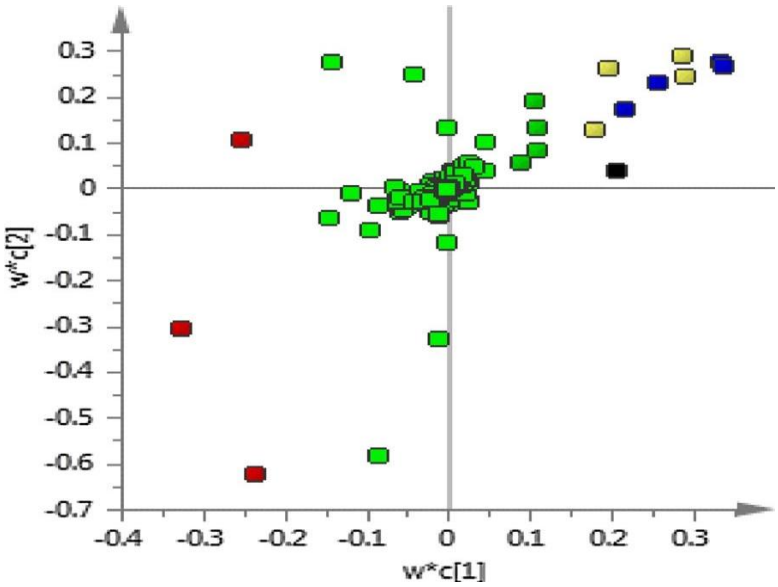

Fig. 3. Loadings Plot of the PLS model applying SNV filter. The first factor (R^2X 20.7%) was represented against the second factor (R^2X 62.3%). The green and black points re- present the variables X and Y, respectively. The variables X (t_R) corresponding to 2-furanmethanol and N-methyl-3-propanediamine are specified by the yellow and blue points, respectively while 1-hidroxy-2-propanone is characterized by the red ones. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

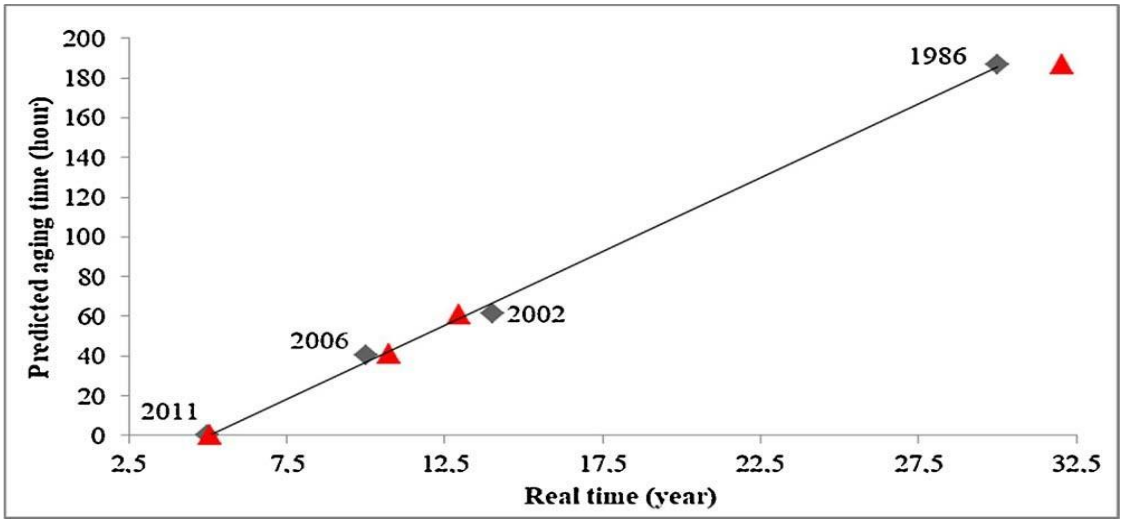

Fig. 4. Graphic of the real time (year) against the predicted aging time (hour) from the PLS model (SNV). Equation Y = 7.43X − 37.25, R^2 0.99. The red triangle represents the estimated date from the regression obtained with the dataset (n = 3) leaving out the test sample. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Finally, an indirect approach based on the identification of characteristic compounds of the period of the document was explored as second approach. In a visual analysis of the pyrograms at 400 °C, a similar profile was observed in all the documents (Fig. 5). In general, at greater aging the intensity of the chromatographic signals was higher as was also observed in the artificial aging study. The main compounds of these pyrograms are listed in Table 4. 4-hydroxybutanoic acid and 2-hydroxy-3,4-dimethyl-2-cyclopenten-1-one were detected in all the documents analysed, except in the sample from 1986. The absence of this type of compounds in the oldest document may be indicative of a change in the manufacture of paper [10,26]. If confirmed, the targeted analysis of these compounds could help to determine the age of the document in an indirect way. In addition, at a pyrolysis temperature of 600 °C, propanoic acid was found in all the oldest documents; 1986, 1991 and 1996. 4-hydroperoxy-1-phenyl-1-cyclohexene, 2-methylbenzofuran and naphthalene, formed by the thermal decomposition of cellulose [37], were also detected in all the sheets pertaining to the document of 1986, but not in the rest of the documents (Table 4).

4. Conclusions

The estimation of the document age based on the analysis of paper applying Py-GC/MS has proven to be a promising approach to be taken into account for forensic purposes. It is a micro destructive approach requiring a small portion of document and no sample treatment prior to the pyrolysis stage.

The direct approach by means of a PLS model after SNV spectral filter obtained acceptable accuracy and precision values of aging estimation for white paper samples. However, this method presented difficulties in finding a universal model for samples with a completely different profile and composition. This direct approach requires that the compared documents were kept under similar conservation conditions. From this direct approach, it was established a relation between the natural and artificial aging, in which 5 h in the aging chamber were equivalent to one natural year under policy custody conditions.

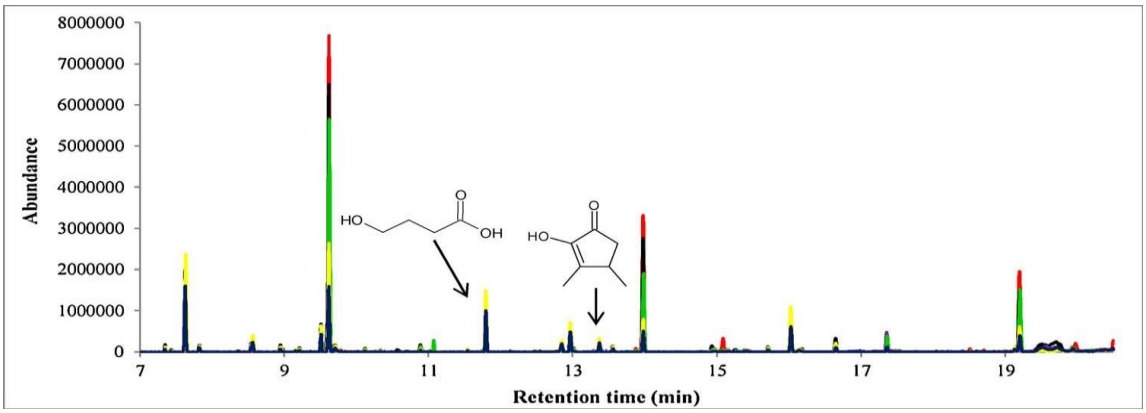

Fig. 5. Pyrograms of the documents under police custody; 2011 (blue line), 2006 (yellow line), 2002 (green line), 1996 (purple line), 1991 (black line) and 1986 (red line). Analysis (pyrolytic and chromatographic) conditions as in Fig. 1. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 4. Identification of characteristic compounds associated with certain chromatographic peaks detected in each page of the documents. Matching higher than 90% for NIST 11 mass spectra library.

Compound	t_R (min)	M (g/mol)	Doc 1986	Doc 1991	Doc 1996	Doc 2002	Doc 2006	Doc 2011
Pyrolysis temperature of 400 °C								
4-Hydroxybutanoic acid	11.54	104.11		×	×	×	×	×
2-Hydroxy-3,4-dimethyl-1,2-cyclopenten-1-one	13.21	126.15		×	×	×	×	×
Pyrolysis temperature of 600 °C								
Propanoic acid	10.53	74.08	×	×	×			
4-Hydroperoxy-1-phenyl-1-cyclohexene	11.11	191.25	×					
2-Methyl benzofuran	11.18	132.16	×					
Naphtalene	12.74	128.17	×					

Acknowledgments

The authors would like to thank the Documentoscopy section (Madrid) of the General Commissary of Scientific Police for facilitating the documents under police custody and explain their forensic necessities. Authors also thank to Dr. Martín Alberca from the Research Group in Forensic Chemistry (INQUIFOR) at the University of Alcalá for his useful discussion. Likewise, we thank Dra. Enara Artetxe and Dra. Itxaso Maguregui of the Faculty of Fine Arts of the University of the Basque Country (UPV/EHU) for the technical support with the accelerated aging chamber. The authors would also like to thank the General Services of Research (SGIker) of the UPV/EHU for the technological

support in the development of this work too. Finally, L. Ortiz- Herrero thanks UPV/EHU for the pre-doctoral fellowship.

References

[1] M.R. López-Ramírez, N. Navas, L.R. Rodríguez-Simón, J.C. Otero, E. Manzano, Study of modern artistic materials using combined spectroscopic and chromato- graphic techniques. Case study: painting with the signature Picasso, Anal. Methods 7 (2015) 1499–1508.

[2] M. Calcerrada, C. García-Ruiz, Analysis of questioned documents: a review, Anal. Chim. Acta 853 (2015) 143–166.

[3] M. Ezcurra, J.M.G. Góngora, I. Maguregui, R. Alonso, Analytical methods for dating modern writing instrument inks on paper, Forensic Sci. Int. 197 (2010) 1–20.

[4] M. Manso, M.L. Carvalho, Application of spectroscopic techniques for the study of paper documents: a survey, Spectrochim. Acta Part B At. Spectrosc. 64 (6) (2009) 482–490.

[5] Y. Keheyan, V. Aharonian, Proceedings 4th International Congress on Science and Technology for the Safeguard of Cultural Heritage in the Mediterranean Basin, v. 2, 6–8 December, Cairo Egypt, 2009, pp. 542–551.

[6] Y. Keheyan, Py/GC/MS analyses of historical papers, BioResources 3 (3) (2008) 829–837.

[7] C. Avataneo, M. Sablier, New criteria for the characterization of traditional East Asian papers, Environ. Sci. Pollut. Res. 24 (2017) 2166–2181.

[8] C. Weyermann, B. Spengler, The potential of artificial aging for modelling of natural aging processes of ballpoint ink, Forensic Sci. Int. 180 (2008) 23–31.

[9] B. Han, G. Daheur, M. Sablier, Py-GCxGC/MS in cultural heritage studies: an il-lustration through analytical characterization of traditional East Asian handmade papers, J. Anal. Appl. Pyrolysis 122 (2016) 458–467.

[10] M.C. Area, H. Cheradame, Paper aging and degradation: recent findings and re-search methods, BioResources 6 (4) (2011) 5307–5337.

[11] Y. Keheyan, G. Eliazyan, P. Engel, B. Rittmeier, Py/GC/MS characterisation of naturally and artificially aged inks and papers, J. Anal. Appl. Pyrolysis. 86 (2009) 192-199.

[12] R. Ganzerla, A. Gambaro, E. Cappelletto, M. Fantin, S. Montalbani, M. Orlandi, Characterization of selected paper documents from the archives of Palazzo Ducale (Venice), Italy using various analytical techniques, Microchem. J. 91 (2009) 70–77.

[13] H. Ohtani, T. Komura, N. Sonoda, Y. Taguchi, Evaluation of acidic paper deterioration in library materials by pyrolysis-gas chromatography, J. Anal. Appl. Pyrolysis. 85 (2009) 460–464.

[14] S. Manente, A. Micheluz, R. Ganzerla, G. Ravagnan, A. Gambaro, Chemical and biological characterization of paper: a case study using a proposed methodological approach, Int. Biodeterior. Biodegrad. 74 (2012) 99–108.